Cohort Shepherd II: Verifying Cohort Constraints from Hospital Visits

Travis Goodwin, Kirk Roberts, Bryan Rink, Sanda M. Harabagiu
Human Language Technology Research Institute
University of Texas at Dallas
Richardson TX, 75080
{travis,kirk,bryan,sanda}@hlt.utdallas.edu

Abstract

This paper describes the updated system created by the University of Texas at Dallas for content-based medical record retrieval submitted to the TREC 2012 Medical Records Track. Our system updates our work from the previous year by building a structured query for each cohort that captures the patient's age, gender, hospital status, and medical assertion information. Further, all keywords that encode any medical phenomena from the query are recursively decomposed before being expanded using knowledge from UMLS, SNOMED, Wikipedia, and PubMed co-occurrences. An initial ranking of hospital visits is then obtained using BM25 relevance on an interpolation of these decomposed keywords. Finally, hospital visits are re-ranked according to the constraints extracted in the structured query. Four runs were submitted, comparing pair-wise combinations of complete vs. shallow keyword decomposition and full vs. negation-only assertion processing. Our highest scoring submission achieved an infNDCG score of 0.426.

1 Introduction

As electronic medical records (EMRs) become more ubiquitous throughout the healthcare industry, the necessity of robust, domain-aware information retrieval techniques emerges. In particular, the need to quickly and accurately retrieve medical records corresponding to specific medical constraints – the ability to retrieve patient cohorts cohorts – will be of critical importance as the industry continues to embrace new technologies.

This type of content-based, domain-specific retrieval is precisely what the Text REtrieval Conference (TREC) 2012 medical records track intends to advance. A continuation of the track which began in TREC 2011, participants were provided with a corpus of free-text electronic medical records (EMRs)[1] and a mapping from each EMR to its corresponding hospital visit[2]. Additionally, the thirty-five evaluation queries from TREC 2011 were provided for training or tuning usage. These queries (referred to as *topics* by the task organizers) each target a specific hospital patient cohort, characterized by various medical phenomena, as evidenced in table 1.

The task requires that participants return a ranked list of hospital visits, such that the rank of each hospital visit indicates the degree by which its associated EMRs are relevant[3] to the given query.

The remainder of this text is outlined as follows. Section 2 provides an outline of the improved

[1] The electronic medical records were provided by the University of Pittsburgh BLULab NLP Repository, and are available at `http://www.dbmi.pitt.edu/nlpfront`.

[2] The number of EMRs associated with a hospital visit varies from 1 to 418, with a median of 3.

[3] The decision of how to consider relevancy for the entire set of EMRs associated with a given hospital visit was left to each group. We addressed this by merging all EMRs for each hospital visit into a single document.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE
NOV 2012

2. REPORT TYPE

3. DATES COVERED
00-00-2012 to 00-00-2012

4. TITLE AND SUBTITLE
Cohort Shepherd II: Verifying Cohort Constraints from Hospital Visits

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas at Dallas,Human Language Technology Research Institute,Richardson,TX,75080

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the Twenty-First Text REtrieval Conference (TREC 2012) held in Gaithersburg, Maryland, November 6-9, 2012. The conference was co-sponsored by the National Institute of Standards and Technology (NIST) the Defense Advanced Research Projects Agency (DARPA) and the Advanced Research and Development Activity (ARDA). U.S. Government or Federal Rights License

14. ABSTRACT

This paper describes the updated system created by the University of Texas at Dallas for content-based medical record retrieval submitted to the TREC 2012 Medical Records Track. Our system updates our work from the previous year by building a structured query for each cohort that captures the patient's age, gender hospital status, and medical assertion information. Further, all keywords that encode any medical phenomena from the query are recursively decomposed before being expanded using knowledge from UMLS SNOMED, Wikipedia, and PubMed co-occurrences. An initial ranking of hospital visits is then obtained using BM25 relevance on an interpolation of these decomposed keywords. Finally, hospital visits are re-ranked according to the constraints extracted in the structured query. Four runs were submitted, comparing pair-wise combinations of complete vs. shallow keyword decomposition and full vs. negation-only assertion processing. Our highest scoring submission achieved an infNDCG score of 0.426.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

104	Patients diagnosed with localized prostate cancer and treated with robotic surgery.
106	Patients diagnosed who had positron emission tomography (PET), magnetic resonance imaging (MRI), or computed tomography (CT) for staging or monitoring of cancer.
112	Female patients with breast cancer with mastectomies during admission.
119	Adult patients who presented to the emergency room with with anion gap acidosis secondary to insulin dependent diabetes. [SIC]

Table 1: Examples of queries used from TREC 2011.

COHORT-SHEPHERD system. Next, section 3 and its sub-sections detail the QUERY ANALYSIS module and how we analyse the semantic constraints of a given query. Then, section 4 illustrates the KEYWORD EXTRACTION and decomposition module, while section 5 details the methods of expansion utilized by our KEYWORD EXPANSION module. This is followed by an in-depth discussion of how we retrieve of hospital visits within the RETRIEVAL module as explained in section 6. Section 7 and its associated sub-section analyse our re-ranking modules and how our initial ranked set of hospital visits is iteratively re-ranked to form our final ranking. Finally, section 8 provides evaluation followed by analysis in section 9 and a brief conclusion in section 10.

2 System Architecture

What follows is a brief outline of the architecture of our system (illustrated in figure 1), followed by an in-depth look at each component.

The queries presented in this task convey a variety of complex semantic constraints on the targeted patient cohort. As such, our system begins by first detecting these cohort constraints and representing them in a structured form that the computer understands.

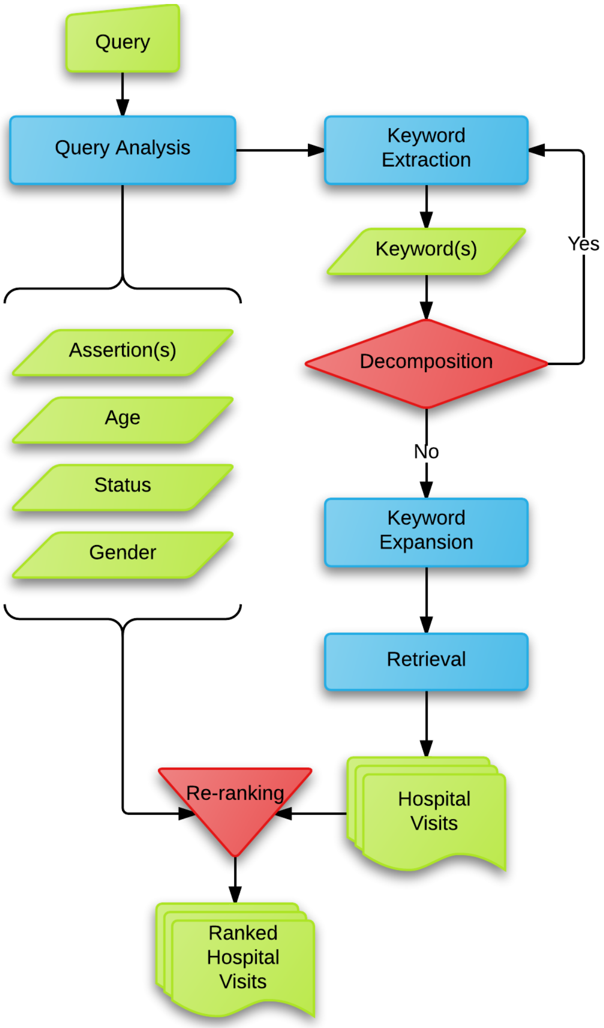

Figure 1: The Architecture of Cohort Shepherd

We accomplish this through the Query Analysis module which consists of several submodules for distilling patient age (e.g. *elderly*, *children*), patient gender(e.g. *women*, male patients), hospital status (e.g. *presenting to the emergency room*, *discharged from the hospital*, *admitted with*), or medical assertion[4] status which captures the existence, absence or uncertainty of medical phenomena (e.g. *without a diagnosis of* x, *family history of* x, *recommended for possible* x). Next, the actual medical phenomena itself is detected by the Keyword Extraction module. In this phase keywords are recognized from the query and recursively decomposed into sub-keywords. Additionally, because of the incredible diversity within the diction used throughout the electronic medical records, each keyword is expanded using the following knowledge sources:

1. The Unified Medical Language System (UMLS) Metathesaurus
2. The English Wikipedia redirect database
3. The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) 2011
4. PubMed Central co-occurrence information

After the queries have been thoroughly dissected, an initial ranked list of hospital visits is acquired using Apache Lucene 4.0's BM25 retrieval implementation. Then, this initial ranked list of hospital visits is re-ranked using the knowledge gleaned from the Query Analysis module to yield the final ranking.

3 Query Analysis

Our Query Analysis module is motived by the thirty-five queries from TREC 2011 as well as the practice queries provided by the National Library of Medicine (NLM)[5]. These queries presented an extraordinary variety of complex constraints on their desired patient cohort that require more semantic reasoning than basic keyword matching. As such, based on our analysis of the 2011 and NLM queries, we created four sub-modules to detect any patient age, medical assertion, hospital status, and patient gender constraints imposed by the query, which are described in the following sub-sections.

3.1 Assertion Detection

The cohorts queried in this task are characterized by medical problems, medical treatments, and medical tests. However, as indicated in table 2, the existence, absence or uncertainty of these medical problems treatments or tests may vary. This information – the belief state of a concept – is known as an assertion.

Query Excerpt	Assertion Value
without a diagnosis of x	ABSENT
with a history of x	HISTORICAL
recommended for possible x	POSSIBLE

Table 2: Examples of detected assertions from TREC queries.

In order to detect this information, we manually annotated the assertion status of 2,349 medical concepts (1,183 problems, 614 tests, 552 treatments) and used a Support Vector Machine (SVM) to trained on these annotations to the classify the status of each concept identified in a given query. Our assertion detection technique follows that described in Roberts and Harabagiu [2011]. We use six-way classification of belief status at the concept level, as well as similar features (both for concept detection and assertion classification). Finally, we also utilize the same methods for feature selection.

3.2 Age Detection

Although somewhat rare, some queries targeted patients characterized by a specific age, or age range (such as query 119 in table 1 which targets adult patients only). Patient age information is detected according to manually created grammar extrapolated

[4] A useful description of medical assertions is provided in Roberts and Harabagiu [2011].

[5] NIST endorsed sixty practice topics generated by the NLM based on a priorities published by the Institute of Medicine (IOM). These queries are available currently for TREC participants only at `http://trec.nist.gov/act_part/tracks/medical/NLM_sample_topics.txt`.

from the sixty practice topics provided by the National Library of Medicine. Our grammar is described in detail in Goodwin et al. [2011] captures queries of the form *patients* ***younger than*** x, *patients* ***at most*** x ***years old***, as well as ranges such as *patients in their* ***thirties to sixties***. We also detect common age ranges based on a lexicon of known phrases, such as *children, elderly, adult* have been manually mapped to their numerical ranges.

3.3 Discovering Hospital Status

Queries such as those in table 3 reveal another trend among the cohorts targeted by the TREC medical records task – hospital status. We observed three such criteria, that occurred frequently throughout the 2011 and NLM practice queries: Admission, Discharge, and Emergency Room. The desired hospital status was detected by comparing the lemmatized query against a small set of simple patterns, given in table 3.

Hospital Status	Example Query	Lexical Patterns
Admission	Patients admitted with a diagnosis of multiple sclerosis	• *admit for* • *admit to the hospital for* • *present to the hospital*
Discharge	Patients being discharged from the hospital on hemodialysis	• *discharge*
Emergency Room	Patients with CAD who presented to the Emergency Department with Acute Coronary Syndrome and were given Plavix	• *Emergency Department* • *ED course* • *emergency room*

Table 3: Detected hospital statuses, example queries, and the patterns that detect them

3.4 Gender Detection

Our inspection of the 2011 and NLM practice queries revealed that some cohorts target specific patient genders. For example, the query 112 from table 1 requires only visits pertaining to female patients. In order to detect this information, we created a high-precision lexicon of words that denote male subjects, and another that denotes female subjects. These lexicons are available in tables 4 and 5 respectively.

man	men	boy	boys	dude
dudes	gentleman	gentlemen	guy	guys
lad	lads	he	him	his
himself	male			

Table 4: Lexicon of male gender words

woman	women	female	females	girl
girls	dudette	dudettes	lady	ladies
gal	gals	lass	lasses	lassie
lassies	she	her	hers	herself

Table 5: Lexicon of female gender words

4 Keyword Extraction

The queries presented in the TREC 2011 and 2012 medical record track target specific patient cohorts: groups of people constrained by specific medical problems, treatments, or tests. As such, we must detect these constraints – which we cast as keywords – and represent them in a machine-readable format. We accomplish this through the Keyword Extraction module.

Because medical phenomena are often represented through multi-token, complex nominal phrases, typical keyword extraction loses the semantics encoded by the syntactic structure of the query. Consider, for example, the major phenomena – keywords – extracted from the queries given in table 1: query 104 contains *localized prostate cancer* and *treated with robotic surgery*; query 106 contains *positron emission tomography*, *magnetic resonance imaging*, *staging*, *monitoring*, and *cancer*; query 112 contains *breast cancer*, *mastectomies*; and query 119 contains *anion gap acidosis* and *insulin dependent diabetes*. But how do we know which token sequences constitute a keyword, and when to decompose a token sequence into separate keywords?

We recursively consider all sub-sequences of tokens from each query and check if that sequence corresponds to an article title in Wikipedia. This allows us to capture virtually any medical concept as well as common abbreviations, misspellings, short-hand, phrasal verbs, noun collocations and synonyms. However, many common phrases and stopwords exist as Wikipedia articles. To combat this, we ensure that any matched sequence occurs less than a threshold, λ[6], within the PubMed Central open access subset of biomedical text[7].

Finally, each keyword extracted is decomposed so that it contains, as sub-keywords, any phrases within it which would themselves satisfy the keyword criteria. For example, the keyword *lower extremity chronic wound* in figure 2 contains the sub-keywords *lower extremity* and *chronic wound*; sub-keyword *lower extremity* thus, contains the sub-sub-keywords *lower* and *extremity*, while sub-keyword *chronic wound* contains sub-sub-keywords *chronic* and *wound*. The purpose of collecting a hierarchy of sub-keywords in this way is so that the relationship between them may be retained when retrieval is performed.

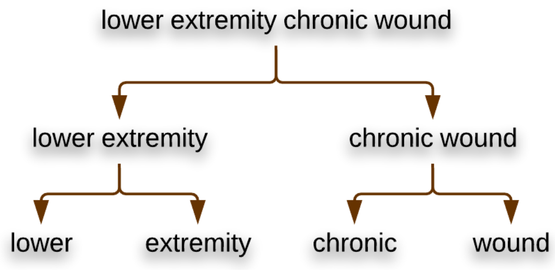

Figure 2: Example of query decomposition for *lower extremity chronic wound.*

5 Keyword Expansion

Within natural language, particularly within medical records, the morphology of words varies extraordinarily both within and between medical texts. To mitigate this diversity of diction, we expand each keyword so as that it may match a variety of lexical forms encompassing synonymy, metonymy, and hyponymy as described in Goodwin et al. [2011]. In order to ease slight variation in syntax, the following simple keyword expansions are performed:

- a WordNet Fellbaum [1998] lemmatized form
- an unabbreviated form based on an internal list of common medical abbreviations
- a form in which all hyphens are padded by spaces
- a form in which all hyphens are replaced by spaces
- a form in which all punctuation is removed

Simple surface form variations are not enough to capture the range of terms doctors use to describe their patients conditions. For example, consider the term *stroke*. This phrase may be referred to as *apoplexy*, *brain attack*, or *cerebrovascular accident*. In order to capture this degree of synonymy, we utilize the Unified Medical Language System (UMLS) Metathesaurus Schuyler et al. [1993]. The UMLS Metathesaurus is a medical ontology which aggregates knowledge from RxNorm, MeSH, SNOMED and other sources. We utilize this knowledge by expanding a given keyword so that it also matches all lexical forms which map to the same CONCEPT ID within the UMLS Metathesaurus database.

Despite the high precision achieved by incorporating knowledge from UMLS Metathesaurus, the recall was not sufficient for our needs. The terms used in the electronic medical records contained spelling variations and a wide variety of slang or less precise synonymy than UMLS encodes. To bridge this knowledge gap, we leveraged the English version of Wikipedia. We used a list of all redirect articles – pages that send the reader to a new article rather rather than containing information on their own. These redirect articles suite

[6] In our case, $\lambda = 30,000$. This was based on observed occurrences of keywords from the TREC 2011 queries.

[7] It is out belief that by using a biomedical corpus, we can more accurately target domain-specific keywords and filter domain-specific stopwords

our needs because they typically correspond to alternate names, spellings, lexical forms, related words, or hyponyms. We use this information by expanding a given keyword such that it corresponds to any lexical forms used as article titles that redirect to the given keyword. For example, using Wikipedia redirects expansions allows us to expand the keyword *hearing loss* to *auditory impairment*, *deaf*, *deafness*, *hard of hearing*, *hearing damage*.

While synonymy and alternations are sufficient for many keyword matches, some questions are constrained by information that requires greater reasoning. Consider, for example, the keyword, *atypical antipsychotics*. Doctors will not use this phrase as-is in their records, but rather, will use hypernyms or metonyms – specific types of atypical antipsychotics in its place. In order to match this kind of variation, we incorporation the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). SNOMED CT is an ontology of clinical terms maintained by the International Health Terminology Standards Development Organisation which documents both clinical terms and, more importantly, the relationships between them. We incorporate this knowledge by expanding a given keyword so as to match any lexical form encoded in SNOMED CT that partakes in the child side of an IS_A, PART_OF, or COMPONENT relationship. By doing so, the keyword *atypical antipsychotics* may be expanded to include *abilify*, *aripiprazole*, *asenapine*, *clozapine*, *clozaril*.

While the previous keyword expansion techniques are sufficient for most scenarios, the text of electronic medical records is often terse, disjoint, and ungrammatical. Additionally, some keywords may require more domain knowledge than what we are able to simulate with mere keyword expansion. As a fall-back, to help mitigate this domain knowledge rift, we expand keywords so that they correspond to related terms. We calculate these related terms using co-occurrence information gleaned from the PubMed Central Open Access Subset (PMC), a collection of freely available biomedical texts. Related was determined by considering the normalized Google distance Cilibrasi and Vitanyi [2007]. The $\mathrm{NGD}(x, y)$ is defined as:

$$\frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log M - \min\{\log f(x), \log f(y)\}}$$

where M is the total number of documents in PMC; $f(x)$ and $f(y)$ are the number of documents containing terms x and y, respectively; and $f(x, y)$ is the number of documents in which x and y co-occur.

We selected the top twenty expansions of sufficient similarity[8] as the expansions for each keyword. For example, *atypical antipsychotics* acquired *olanzapine*, *risperidone*, *quetiapine*, *clozapine*, and *antipsychotic drug*.

6 Hospital Visit Retrieval

After extracting and expanding the keywords that characterize a patient cohort, we must retrieve all relevant hospital visits that match the extracted keywords. This task is accomplished through the use of Apache Lucene 4.0 Hatcher and Gospodnetic [2005].

Prior to retrieval, we created an index over all hospital visits by merging all the electronic medical records associated with each hospital visit into a single document. The various fields encoded in each EMR were retained when indexed (admit diagnosis, chief complaint, etc) so that per-field weights could be adjusted.

For retrieval, each query is represented as an interpolation of its decomposed keywords and their weighted expansions:

$$\begin{aligned}\mathrm{query}(k, \lambda) = \lambda[k + &\alpha\,\mathrm{UMLS}(k)\\ &+\beta\,\mathrm{Wikipedia}(k)\\ &+\gamma\,\mathrm{SNOMED}(k)\\ &+\delta\,\text{Co-Occurrence}(k)]\\ &+ \sum_{s \in \text{sub-keywords}(k)} \mathrm{query}(s, \mu\lambda)\end{aligned}$$

where λ is the initial keyword score; α, β, γ and δ, are the weights associated with the respective keyword expansion method; and μ is the discounting factor $0 <$

[8] See Goodwin et al. [2011] for more information regarding co-occurrence keyword expansion.

$\mu < 1$ (we used $\lambda = 16, \alpha = 12, \beta = 10, \gamma = 8, \delta = 1$, and $\mu = 0.5$). The queries were represented in Lucene using nested SpanNear and SpanOr queries, and relevance was judged using the Okapi BM25 ranking function Robertson and Walker [1994]. This yields a ranked list of hospital visits, ordered by BM25's interpretation of our query representation.

7 Re-ranking

Although our initial Lucene retrieval performs reasonably well for the purpose of ranking documents strictly within respect to keyword relevancy, the queries presented in TREC are characterized by more complex constraints. We address these additional cohort constrains by an iterative re-ranking process: for each constraint identified by the QUERY ANALYSIS module (patient age, patient gender, hospital status, medical assertion value), we heuristically re-rank all hospital visits for a given question. After each constraint has been considered, the final ranking of patient hospital visits is returned as the solution of our system. What follows is a description of each heuristic re-ranking sub-module.

7.1 Re-Ranking by the Patient's Age

The current ranked list of hospital visits are re-ranked with respect to patient age by comparing the frequencies of de-identified patient age information within all the reports associated with each hospital visit. Any hospital visit wherein the number of de-identified age mentions falling outside the numerical range identified by the QUERY ANALYSIS module (described in section 3.2) has its score lowered by 100 where a hospital visit's score is based on the BM25 score described in section 6; any hospital visit lacking any age information has its score lowered by 50 (so that hospital visits that match the desired criteria are elevated to the top).

7.2 Re-Ranking by the Patient's Gender

When considering the patient's gender, we utilize the same lexicons described in section 3.4 and compare the frequency of MALE to FEMALE words in all EMRs associated with a given hospital visit. Hospital visits for which there are more mentions of the opposite gender across all associated EMRs have their current score lowered by 100 (where score is the BM25 score detailed in section 6).

7.3 Re-Ranking according to Assertion Information

As described in section 3.1, each keyword in a given query is associated with a given medical assertion. For the purposes of re-ranking, we attempted to ascertain the degree to which mentions of a given keyword correctly indicate an actually present medical condition, treatment, or test as opposed to an absent or unsure mention. To accomplish this, we assigned the following *negativity value* to each possible assertion value: ABSENT = 1.0, ASSOCIATED_WITH_SOMEONE_ELSE = 1.0, CONDITIONAL = 0.333, CONDUCTED = 0, HISTORICAL = 0.5, HYPOTHETICAL = 0.333, ONGOING = 0, ORDERED = 0, POSSIBLE = 0.333, PRESCRIBED = 0, PRESENT = 0, and SUGGESTED = 0.333. For each keyword mention in all EMRs associated with a given hospital visit, we calculate the sum of the heuristic value. If this sum is more than one-third of the frequency of keyword mentions for a given keyword, we subtract 400 from the current score (where score is the BM25 score illustrated in section 6) such that if multiple keywords satisfy this criteria the score may be lowered multiple times.

7.4 Re-Ranking based on the Patient's Hospital Status

The goal of the hospital status re-ranker is to promote hospital visits wherein at least one EMR that matches a keyword also satisfies the requirements of the patients hospital status detected in section 3.3. In order to achieve this, we consider the meta-data associated with each EMR (the *type* and *subtype* fields which indicate the type of each electronic medical report), as well as context for each keyword match: the previous section header, based on a simple section detection algorithm that looks for the last fully capitalized sentence ending with a colon (e.g. *DISCHARGE SUMMARY:*), and the lemmatized sentence containing the given keyword. For example, when detecting hospital visits that satisfy patient admission criteria, we look for EMRs that have the *subtype* of ADMISSION, or keywords that fall within a section whose header contains *ADMISSION* or *ADMITTING* or whose lemmatized sentence contains *admit for*, *admit to the hospital for*, or *present to the hospital*. Likewise, the criteria for detecting patients discharged from the hospital is an EMR with the *type* of DS or *subtype* or DISCHARGE or any sentence containing the lemma *discharge* used as a verb. Finally, the requirements for asserting EMRs pertaining to the emergency room involves checking if the EMR's *type* is ER, if any keyword's section header contains *EMERGENCY DEPARTMENT* or *ED*, or if any keyword match lies within a lemmatized sentence containing *Emergency Department*, *ED course*, or *emergency room*. Visits wherein at least one EMR did not satisfy the requirements of any detected patient hospital status constraints have their score lowered by 50.

8 Performance Evaluation

We provided four submissions to NIST for TREC 2012. UTDHLTA represents our system as described with no modifications, UTDHLTNA represents our system wherein only strict negations are considered when re-ranking for assertions (that is, only ABSENT or ASSOCIATED_WITH_SOMEONE_ELSE). UTDHLTASK denotes a modification to our system wherein during keyword extraction, keywords may only be decomposed once (shallow decomposition). UTDHLTNASK denotes a version of our system wherein both strict negations and shallow keyword decomposition are enforced. Table 6 summarizes our results for TREC 2012.

Submission	iAP	iNDCG	BPref	P10
UTDHLTA	0.203	0.425	0.311	0.4213
UTDHLTASK	0.188	0.410	0.337	0.4553
UTDHLTNA	0.206	0.426	0.336	0.4532
UTDHLTNASK	0.199	0.424	0.343	0.4489

Table 6: Performance evaluations for TREC 2012. iAP refers to the inferred average precision; iNDCG refers to the inferred normalized discounted cumulative gain, BPref refers to the binary preference; and P10 refers to the precision of the first ten results.

Submission	iAP	iNDCG	iP10
NONE	0.2041	0.4246	0.4447
+AGE	0.2044	0.4248	0.4447
+GENDER	0.2044	0.4248	0.4447
+STATUS	0.2041	0.4249	0.4447
+ASSERTION	0.2139	0.4538	0.4652

Table 7: Re-ranking experiments: effects of adding in each re-ranking component to a baseline system of no re-ranking.

9 Discussion

It is clear from this table that UTDHLTNA was our best performing submission according to the primary metric of this task, inferred average precision. UTDHLTA closely follows it in performance, and the performance difference is small enough that, presumably, the naïve heuristic is at fault. Table 7 shows the impact of adding each re-ranking method of our re-ranking module to the overall score. It is to be noted that the results obtained when all re-ranking methods have been applied are different than the results of the UTDHLTA system submitted, which is due to small bug fixes. However, it is clear that incorporating assertion information yielded significant improvement in ranking.

Assertions encode an incredible amount of semantics from the underlying text, and properly utilizing this information could be of great value to any retrieval system capable of utilizing this knowledge. Future work would benefit from learning weights and using a principled approach to incorporating assertion information into their information retrieval systems.

10 Conclusion

The 2012 Text REtrieval conference marks the second year of the Medical Records Track which begun in 2011. This track sponsors the task of retrieving ranked electronic medical records, grouped by patients' hospital visit, and ranked according to the visit's relevance to a given query. The queries in this task targeted patient cohorts, typically characterized by specific medical treatments, conditions, or tests, as well as specific patient constraints.

We approached this task by extracting the constraints encoded by a given cohort (patient's age, patient's gender, patient's hospitalization status, and keyword assertion status) and the keywords that encode any medical phenomena found in the query. These keywords were recursively decomposed and then expanded using knowledge from UMLS, SNOMED, and Wikipedia, as well as PubMed Central co-occurrence information. We then perform retrieval to achieve an initial ranking of hospital visits (based on a BM25 relevance model of an interpolation of all decomposed keywords and their associated expansions). Finally, using the constraints extracted earlier, we iteratively re-rank the set of hospital visits for each constraint until we achieve our final ranking.

The incorporation of assertion status (existence, absence or uncertainty) of a medical condition represents an important step towards truly understanding the semantics behind what is actually being said in a given electronic medical record and could play an integral role in future retrieval systems working in the medical domain. Utilizing this knowledge, however, leaves much room for future work as the complexity involved is nontrivial.

11 Acknowledgements

We are indebted to Dr. Michael Skinner, M.D. a Professor of Surgery at UT Southwestern Medical Center as well as pediatric and general surgeon, for all of his invaluable clinical information.

References

R.L. Cilibrasi and P.M.B. Vitanyi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, pages 370–383, 2007.

C. Fellbaum. *WordNet: An Electronic Lexical Database*. The MIT press, 1998.

T. Goodwin, B. Rink, K. Roberts, and S.M. Harabagiu. Cohort shepherd: Discovering cohort traits from hospital visits. In *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, 2011.

E. Hatcher and O. Gospodnetic. *Lucene in Action*. Manning Publications, 2005.

K. Roberts and S.M. Harabagiu. A flexible framework for deriving assertions from electronic medical records. *Journal of the American Medical Informatics Association*, 18(5):568–573, 2011.

S.E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 232–241. Springer-Verlag New York, Inc., 1994.

P.L. Schuyler, W.T. Hole, M.S. Tuttle, and D.D. Sherertz. The umls metathesaurus: Representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217, 1993.